

Multiple-spell Regression Models for Duration Data

By ALFRED HAMERLE†

University of Konstanz, West Germany

[Received January 1987. Final revision February 1988]

SUMMARY

General models for multiple-spell duration data are considered. A general theory which indicates how the successive spells of an individual are generated by an underlying stochastic process is presented. Various special cases of the general model are discussed. The implications of different timescales are investigated: different timescales lead to different underlying stochastic processes such as Markov processes or semi-Markov processes. Occasionally common computer programs for duration data such as SAS, BMDP, GLIM and RATE can be used without further programming. Finally, multiple-spell models are applied to the duration of unemployment, and analysing the data in a single-spell framework is shown to lead to false interpretations and conclusions.

Keywords: Completely and partially parameterized models; Different timescales; Maximum likelihood estimation; Multistate–multiepisode duration model; Transition-specific hazard rate

1. Introduction

Event history or failure time data are collected in follow-up studies, retrospective studies and sometimes in longitudinal panels. The data record qualitative changes over time in some important variables, e.g. job changes, technical defects and deaths. The main purpose of the statistical analysis is to investigate the time before a certain event occurs. It is important to evaluate the association of exposure, treatment and prognostic factors with the distribution of time until the event occurs, and, by analogy with conventional regression analysis, much of the attention in duration analysis focuses on the effect of covariates on durations. For the statistical analysis of such dynamic processes hazard rate models can be used where the hazard rate depends on independent variables. The statistical theory of duration data is described by Kalbfleisch and Prentice (1980), Lawless (1982), Cox and Oakes (1984), Tuma and Hannan (1984), Blossfeld *et al.* (1986, 1988) and others.

Most of the methods in the literature describe the statistical analysis of single-spell duration data, because most of the regression models for durations were developed in biostatistics by analysing survival times of patients in a clinical trial. In other areas such as economics, sociology, ecology, psychology or industrial engineering, study subjects can experience more than one event or failure as time elapses and, moreover, these events or failures may be of various kinds. Most of the available software

†Address for correspondence: Department of Statistics, University of Konstanz, Postfach 5560, D-7750 Konstanz, West Germany.

(e.g. BMDP, SAS, GLIM and RATE) was originally made for single-spell models. In reducing the analysis of multiple-spell data to the analysis of single-spell models, some restrictive assumptions are usually made such as that successive episodes of an individual are independent (e.g. Arminger (1984)), that the effects of the explanatory variables are the same in all episodes or that the duration distributions only depend on time since entry into the present state, but when invoking such assumptions much information about the underlying dynamic process is lost, and ignoring important aspects of this process may lead to false conclusions.

Several researchers (Kalbfleisch and Prentice, 1980, ch. 7; Prentice *et al.*, 1981; Wu, 1982; Kay, 1983; Heckman and Singer, 1984) have studied models for multiple-spell duration data without these restrictive assumptions. Most of the papers consider special cases, in particular semi-Markov models. The purpose of this paper is three-fold: firstly we shall present a general theory which indicates how the successive spells of an individual are generated by an underlying stochastic process which describes the individual behaviour over time. There are two related processes, the transition process and the duration process. The link between the two processes can be expressed in terms of transition-specific hazard or exit rates depending on covariates. Various special cases of the general model are discussed. Secondly, we shall investigate the implications of different timescales. Different timescales lead to different underlying stochastic processes, for example to inhomogeneous Markov processes or semi-Markov processes. Thirdly, we shall deal with estimation procedures. Full and partial likelihood methods are investigated and the applicability of available computer programs discussed. An application of the models to the duration of unemployment is given in the last section.

2. Statistical Model

2.1. Underlying Stochastic Process

In the statistical model for the underlying stochastic process points of time in which transitions occur are represented by a series of non-negative random variables $0 = T_0 \leq T_1 \leq \dots$ and the state variable is characterized by a series of random variables $\{Y_k: k = 0, 1, 2, \dots\}$, $Y_k \in \{1, \dots, m\}$. The corresponding stochastic process which describes the state of an individual at time t is $Z(t)$, $t \geq 0$, where

$$Z(t) = Y_{k-1} \quad \text{for} \quad T_{k-1} \leq t < T_k, \quad k = 1, 2, \dots$$

$Z(t)$ is a continuous time, discrete state process which can be decomposed into two related processes, the transition process and the duration process. The transition process is the stochastic process which governs the transitions between states. The duration process governs the length of stay in a particular state. Let us suppose, then, that for every individual and episode k a vector x_k of relevant covariates is measured. The vector of covariates may contain metric or dummy variables or both. For the moment we assume that the covariates are time invariant, but most of the methods can be generalized to allow for time-dependent covariates. If an individual is in the k th episode, his previous history of the process until t_{k-1} is collected in H_{k-1} , i.e. $H_{k-1} = \{y_0, t_1, y_1, x_1, \dots, t_{k-1}, y_{k-1}, x_{k-1}\}$. Both the duration and the transition process can be simultaneously characterized by a transition-specific or cause-specific hazard rate

$$\lambda_j^k(t|x_k, H_{k-1}) = \lim_{\Delta t \rightarrow 0} P(t \leq T_k < t + \Delta t, \quad Y_k = j | T_k \geq t, x_k, H_{k-1}) / \Delta t, \\ t \geq t_{k-1}. \quad (2.1)$$

$\lambda_j^k(t|x_k, H_{k-1})$ will be identically zero for $t < t_{k-1}$. The possibility that some states are not attainable from certain states or that the set of attainable states varies with k in a more general manner can be accommodated in equation (2.1) by restricting equation (2.1) to be identically zero for the appropriate (k, j) values.

The total hazard rate of the k th episode is

$$\lambda^k(t|x_k, H_{k-1}) = \sum_j \lambda_j^k(t|x_k, H_{k-1}). \quad (2.2)$$

The survivor function of the k th episode is given by

$$S^k(t|x_k, H_{k-1}) = \exp\left[-\int_{t_{k-1}}^t \lambda^k(u|x_k, H_{k-1}) du\right]. \quad (2.3)$$

Because $\lambda^k(t|x_k, H_{k-1})$ is identically zero for $t < t_{k-1}$, the survivor function $S^k(t|x_k, H_{k-1}) = 1$ for $t < t_{k-1}$.

Finally we can define a (sub)density function for transition to state j and episode k , given x_k and H_{k-1}

$$f_j^k(t|x_k, H_{k-1}) = \lambda_j^k(t|x_k, H_{k-1})S^k(t|x_k, H_{k-1}). \quad (2.4)$$

Sometimes it is useful to express the survivor function in terms of the transition-specific hazard rates: from equation (2.2)

$$S^k(t|x_k, H_{k-1}) = \prod_{j=1}^m \exp\left[-\int_{t_{k-1}}^t \lambda_j^k(u|x_k, H_{k-1}) du\right]. \quad (2.5)$$

We assume that the beginning of the process is identical with the beginning of the observation period so that there are no left-censored spells. This could be restrictive in economic and sociological panels because here data are random samples of interrupted spells or are spells that begin after the start date of the sample (see, for example, Ridder (1984), Heckman and Singer (1984) or Hamerle (1987)).

2.2. Specification of Hazard or Exit Rate

From equations (2.4) and (2.5) it becomes apparent that the probability law of the transition process and the duration process can be completely expressed in terms of transition-specific hazard or exit rates. Therefore, transition-specific hazard rates are the basic concept for the construction of regression models for duration data. If we specify the transition-specific hazard rates, all relevant duration distributions and transition probabilities can be calculated.

For a complete specification of the hazard rate we must fix

- (a) the time dependence of the hazard rate,
- (b) how much of the previous history is included and
- (c) the dependence on the incorporated covariates.

Let vector $z_k, k = 1, 2, \dots$, be x_k , the covariates measured at the beginning of the k th episode, and the relevant elements of H_{k-1} . The dimension of z_k is p_k , and the transition-specific hazard rate of the k th spell or episode is now denoted by $\lambda_j^k(t|z_k)$. We only distinguish two alternatives. Firstly t is the time since the beginning of the process, e.g. an individual's age, the time elapsed since the diagnosis of a certain disease or in general time since the occurrence of a specific event. Secondly, assuming

that the hazard rate depends on the time since the last transition,

$$\lambda_j^k(t|z_k) = \lambda_j^k(t - t_{k-1}|z_k), \quad (2.6)$$

these models are suitable when the hazard rates for the duration process can be more parsimoniously expressed in terms of gaps between events rather than in terms of total observation time (Heckman and Singer, 1984). Semi-Markov processes or renewal processes are special cases of equation (2.6).

If we define parametric models for the transition-specific hazard rate, it is assumed that the hazard rate is known except for a finite number of parameters. We assume that there is a parameter vector θ which determines time dependence and, in addition, we assume that a suitable dimensioned parameter vector β describes the effects of the covariates. In general, the parameters θ and β may depend on the origin state y_{k-1} , the destination state y_k and the number k of the current episode, i.e. θ_{ij}^k and β_{ij}^k (given $y_{k-1} = i$ and $y_k = j$), but for the remainder of this section we drop the indices to simplify notation.

The hazard rate can be written as

$$\lambda(t|z) = \lambda(t|z; \theta, \beta). \quad (2.7)$$

The alternative specification $\lambda(t - t_{k-1}|z; \theta, \beta)$ differs from equation (2.7) in the time dependence.

Examples of parametric models are the exponential and the Weibull models. These models are proportional hazards models, and parametric forms of the Cox model $\lambda(t|z) = \lambda_0(t) \exp(z'\beta)$. An example for a non-proportional hazards model is the log-logistic model which is used in Section 4.

Many restrictions can be placed on the function $\lambda_0(t)$ or on the parameters θ and β . In the general model $\lambda_0(t)$ as well as θ and β may depend on the origin state, the destination state and the serial order of the episode. We shall focus on two classes of restrictions: restrictions on time dependence and on the covariate effects. We consider a specific (i, j) transition. Reintroducing the subscript k for the serial order of the episode, the base-line hazard rate is denoted by $\lambda_{0k}(t)$ and the parameter vectors by θ_k and β_k . Again we only give the formulae for models with time parameter t , but models depending on $t - t_{k-1}$ can be derived analogously.

In the first model we assume that the parameters θ_k determining time dependence are the same over episodes, i.e. $\theta_1 = \theta_2 = \dots = \theta$; for the Cox model the base-line hazards are restricted to $\lambda_{01}(t) = \lambda_{02}(t) = \dots = \lambda_0(t)$.

In the second model we also restrict the corresponding regression coefficients to be equal over episodes, i.e. $\beta_1 = \beta_2 = \dots = \beta$ for those covariates which are measured in each episode. For simplicity we assume that the same covariates are measured in each episode. Otherwise, the model has to be slightly modified. Cox models with these restrictions were proposed by Kay (1983) in this context.

2.3. Estimation

A complete event history of individual i over some observation time span requires the following data information: the initial state y_{i0} , the number n_i of spells in the observation period, the points in time t_{i1}, \dots, t_{in_i} in which state transitions occur, states y_{i1}, \dots, y_{in_i} , corresponding to the state transitions, an indicator δ_i , capable of distinguishing whether or not the n_i th spell is censored, and covariate vectors x_{i1}, \dots, x_{in_i} .

The derivation of the likelihood function for the general multistate–multiepisode case is given in Appendix A. Here we restrict ourselves to a specific (l, j) transition and suppress the subscripts l and j to simplify the notation. The following arguments hold for each admissible (l, j) transition. The likelihood expression which we are considering subsequently is

$$L = \prod_k \prod_{i=1}^N \lambda^k(t_{ik}|z_{ik})^{\delta_{ik}} \left\{ \exp \left[- \int_{t_{i,k-1}}^{t_{ik}} \lambda^k(u|z_{ik}) du \right] \right\}^{\varepsilon_{ik}}, \quad (2.8)$$

where

$$\delta_{ik} = \begin{cases} 1 & \text{if the } k\text{th episode of individual } i \text{ is terminated at time } t_{ik} \\ & \text{by a transition} \\ 0 & \text{otherwise} \end{cases}$$

and

$$\varepsilon_{ik} = \begin{cases} 1 & \text{if individual } i \text{ experiences at least } k \text{ episodes} \\ 0 & \text{otherwise.} \end{cases}$$

2.4. Estimation of Parametric Models

The likelihood expression (2.8) can be used to estimate fully parameterized models of type (2.7). For this the assumed parametric form of the hazard rate is inserted in expression (2.8). If the parameter vectors (θ_k, β_k) , $k = 1, 2, \dots$, have no elements in common, expression (2.8) can be maximized separately for each k , but only for models

$$\lambda^k(t|z_k) = \lambda^k(t - t_{k-1}|z_k; \theta, \beta),$$

where the time parameter is time since entry into the current state, can we use common computer programs for parametric single-spell models, e.g. the procedure LIFEREG of SAS. If the time parameter is observation time, these programs are not applicable.

The use of programs for single-spell models to estimate general multiple-spell models where time dependence as well as the regression coefficients depend on the number of the episode has a further limitation. If some of the elements of the vectors θ_k and β_k are equal across episodes, the episode-specific factors in expression (2.8) cannot be maximized separately. The parameters have to be estimated simultaneously and this is in general not possible with the available programs. The simultaneous estimation of all parameters can be done with the available programs if we restrict ourselves to models where we assume that the parameters determining the time dependence are equal across episodes whereas the regression coefficients β_k may vary from episode to episode.

The hazard rate of the k th episode is then

$$\lambda^k(t|z_k) = \lambda^k(t|z_k; \beta_k, \theta).$$

For estimation of hazard rate models of this type with programs for single-spell models a rearrangement of regression vectors and parameter vectors is necessary. Let K be the maximum number of episodes, and let \tilde{x}_i denote the covariates which are the same across episodes, e.g. sex, race and demographic variables which remain fixed

during the observation period. The covariates of individual i which vary over episodes are denoted by x_{i1}, \dots, x_{in_i} , $1 \leq n_i \leq K$. The dimension of \tilde{x}_i is q ; the dimension of x_{ik} is q_k . Now for each individual we define new design vectors, again denoted by z_{ik} ,

$$z'_{i1} = (\tilde{x}'_i, x'_{i1}, 0, \dots, 0), \dots, z'_{in_i} = (\tilde{x}'_i, 0, \dots, 0, x'_{in_i}, 0, \dots, 0),$$

$$1 \leq n_i \leq K.$$

The dimension of the new design vector z_{ik} is

$$q + \sum_{k=1}^K q_k.$$

Corresponding to the design vectors a suitable dimensioned parameter vector $\beta' = (\tilde{\beta}', \beta'_1, \dots, \beta'_K)$ is defined. Then, we have $z'_{ik}\beta = \tilde{x}'_i\tilde{\beta} + x'_{ik}\beta_k$. With this rearrangement we can use the data of all episodes and maximum likelihood estimation based on the likelihood expression (2.8) can be done with a program for single-spell models. The successive spells of an individual are taken as though they were episodes from different individuals with the corresponding covariate values. Because of the construction of z_{ik} we always obtain the correct part of the likelihood (2.8). In this restricted model tests for equality of some elements of β_k across episodes can easily be performed using the likelihood ratio statistic, the Wald statistic or the score statistic.

Finally we discuss the model with the further restriction that corresponding regression coefficients are also equal across episodes. Since this model is a special case of the preceding model, it can be estimated in the same manner. However, because of the further restriction of equal regression coefficients over episodes the procedure of constructing new design vectors as described is not necessary. The n_i spells of individual i can be treated as though they stem from n_i different individuals with the corresponding covariate values, and all $\sum_{i=1}^N n_i$ spells are analysed together with a program for single-spell models, yielding maximum likelihood estimates for θ and β .

2.5. Estimation of Regression Coefficients in Cox Models

The proportional hazards model of Cox contains an unspecified and unknown base-line hazard rate. In the most general case the base-line hazard as well as the regression coefficients connected with the covariates may depend on the origin state, the destination state and the serial order of the spell.

Consider now inference on the relationship between the transition- and episode-specific hazard functions and the covariates. Cox (1972, 1975) proposed a partial likelihood approach for the estimation of the regression coefficients. We shall not give a detailed derivation of the partial likelihood here. See instead, for example, Kalbfleisch and Prentice (1980), ch. 4 and 5, or Wong (1986). For our general multiple-spell-multiple-state hazard rate model a partial likelihood may be derived as

$$PL = \prod_{l,j} \prod_k \prod_{i=1}^{d_{lj}^k} \frac{\exp(x_{lji}^{k'} \beta_{lj}^k)}{\sum_{r \in R_{jk}(t_{lj}^k)} \exp(x_r^{k'} \beta_{lj}^k)} \quad (2.9)$$

where $t_{lj1}^k, t_{lj2}^k, \dots$ are the d_{lj}^k uncensored durations in which the k th episode started in state l and is terminated by a transition to state j ($l \neq j$), and where x_{lji}^k is the

covariate vector belonging to the individual with duration t_{lji}^k . The risk set $R_{lk}(t)$ is the set of individuals in state l that are at risk of their k th transition just before time t ($k - 1$ but not k transitions have occurred before t). The (l, j, k, i) contribution to equation (2.9) derives from the probability that a transition occurs on the study subject with covariate vector x_{lji}^k at time t_{lji}^k , given the risk set $R_{lk}(t_{lji}^k)$ and given the occurrence of the k th transition on a study subject at time t_{lji}^k , the transition being a transition to state j .

Like the full likelihood (see Appendix A) the partial likelihood (2.9) again factors into a separate component for each l and j which can be maximized regarding a specific transition from state l to state j . Therefore, we again consider a specific (l, j) transition dropping the indices l and j to simplify the notation. First we consider the multiepisode model with episode-changing regression coefficients and episode-changing base-line hazards

$$\lambda^k(t|z_k) = \lambda_{0k}(t) \exp(z_k' \beta_k)$$

where z_k again contains x_k and relevant elements of the previous history of the process.

An appropriate partial likelihood for this model is

$$PL_1 = \prod_k \prod_{i=1}^{d_k} \frac{\exp(z_{ik}' \beta_k)}{\sum_{r \in R_k(t_{ik})} \exp(z_{rk}' \beta_k)} \quad (2.10)$$

where t_{1k}, t_{2k}, \dots are the d_k uncensored durations of the k th episode and z_{ik} is the covariate vector of the individual with duration t_{ik} . The risk set $R_k(t)$ is the set of individuals at risk of their k th event at time t .

If the parameter vectors β_k have no elements in common, equation (2.10) can be maximized separately for each k . For this common computer programs for single-spell models can be used without difficulty, if it is assumed that the base-line hazard rates only depend on time since the last transition, i.e. $\lambda_{0k}(t) = \lambda_{0k}(t - t_{k-1})$. Usually, in single-spell models the starting point of the spell is set to zero. Then the risk set decreases as time elapses. This also holds in multiepisode models, provided that time is measured as waiting time in the current state, i.e. the clock is reset to zero after each transition.

If the timescale represents observation time or time since the occurrence of an individual specific event such as birth or entry into the labour market, the situation is more complicated. In this case, all subsequent starting and ending times of a person's episodes are calculated as the time since the occurrence of this event. Then, for the k th episode, $k \geq 2$, an individual only belongs to the risk set $R_k(t)$ if he is at risk of his k th event at time t , i.e. $k - 1$ but not k events have occurred before time t . Here the risk set may also increase as time goes on. The reason is that in a certain time interval some individuals can have their $(k - 1)$ th event and are then at risk of their k th event, but no individual has his k th event in this time interval. The risk set is increasing in this time interval. A further limitation of the application of the general model arises if some of the components of the parameter vectors β_k are equal across spells. Further programming is needed if we want to test the hypothesis $H_0: \beta_1 = \beta_2 = \dots$

For the second model the episode k , $k = 1, 2, \dots$, has a common shape function and the hazard function can be written as $\lambda^k(t|z_k) = \lambda_0(t) \exp(z_k' \beta_k)$. An appropriate

partial likelihood for this model is (Kalbfleisch and Prentice (1980), p. 184)

$$PL_2 = \prod_k \prod_{i=1}^{d_k} \frac{\exp(z'_{ik} \beta_k)}{\sum_m \sum_{r \in R_m(t_{ik})} \exp(z'_{rm} \beta_m)}. \quad (2.11)$$

Since the denominator of equation (2.11) contains all β_k , $k = 1, 2, \dots$, this partial likelihood does not factor into a separate component for each k and the coefficients β_k need to be estimated simultaneously. This can be done just as in the preceding section by defining new design vectors z_{ik} and a new parameter vector. Then, all spells are gathered and analysed with a program for single-spell models. Assuming that the timescale is the time since the last transition, for a given t_{ik} , the denominator of equation (2.11) contains all spells which last at least t_{ik} . There may be several spells for an individual. The corresponding exponential term is chosen according to the serial order of the spell. For this it is sufficient to form only one risk set. The correct exponential term is chosen through the special construction of the design vectors z_{ik} . For the timescale representing observation time the procedure is similar.

In the third proportional hazards model that we have investigated it is assumed that in addition to the shape functions the regression coefficients are also the same over spells. From equation (2.11) we have (for the model depending on $t - t_{k-1}$)

$$PL_3 = \prod_k \prod_{i=1}^{d_k} \frac{\exp(z'_{ik} \beta)}{\sum_m \sum_{r \in R_m(v_{ik})} \exp(z'_{rm} \beta)}. \quad (2.12)$$

where $v_{ik} = t_{ik} - t_{i,k-1}$. However, PL_3 is equivalent to the partial likelihood which arises if a person's episodes are treated as though they stem from different individuals and all episodes are analysed together with a single-spell model. There is no need for the construction of design vectors z_{ik} as described earlier.

Finally we note that the asymptotic properties of the resulting maximum likelihood estimators or maximum partial likelihood estimators have not yet been investigated for all situations. Andersen and Gill (1982) and others have proved consistency and asymptotic normality of the estimator $\hat{\beta}_k$ in Cox models for single-spell data. Borgan (1984) investigated multiplicative parametric hazard rate models and proved consistency and asymptotic normality of the maximum likelihood estimator in the single-episode case. These results can readily be generalized to the multiple-episode case (see, for example, Hamerle (1985)). All these researchers assume that the timescale is the observation time or the time since the well-defined beginning of the process. A certain class of semi-Markov models where the time is reset to zero after each transition has been investigated by Voelkel and Crowley (1984). We know of no rigorous proof of the asymptotic properties of the maximum likelihood estimators for non-proportional hazards models such as the log-logistic model and others.

3. Empirical Analysis

In this section multiple-spell hazard rate models are applied to the duration of unemployment in Bavaria. The analysis is based on data from *Erhebung zur Struktur der Bewegungsvorgänge bei Arbeitslosen in der Zeit vom 26.5. bis 8.6.1983* of the Federal Labour Office of West Germany. There were 7660 spells, of which 5848 are first spells and 1812 second spells of unemployment. There are no censored spells, and

the covariates are individual characteristics. First we carried out exploratory analysis and applied life-table methods to the data. Dividing the sample into different strata according to some covariates, the hazard rate estimates showed a typical shape: for all groups it firstly increased and then decreased. Therefore we chose the log-logistic model. Its hazard rate is given by

$$\lambda^k(t|z_k) = \frac{\exp(z'_k \beta_k) \theta_k t^{\theta_k - 1}}{1 + \exp(z'_k \beta_k) t^{\theta_k}}.$$

We only investigate return to employment and in particular transitions to employment where the jobs are found by the individuals themselves. Arrangements by the Federal Labour Office define another state and transitions to this state are not analysed here.

For the present analysis the three model specifications described in the previous section have been used. The time parameter is always time since entry into the last period of unemployment. This leads to models of type (2.6). The first model corresponds to the general model (2.7) where the parameters which determine time dependence as well as the regression coefficients depend on the serial order of the spell of unemployment. In the second model time dependence of the exit rate is assumed to be the same over consecutive episodes. The third model imposes the additional restriction that the regression coefficients are constant across episodes. In most applications an approach based on the third model is used, but our analysis shows that the results based on such an approach which ignores the dynamic aspects of the process, in particular the serial order of the employment-unemployment history, may be misleading.

The calculations were done with the P3RFUN program written by Petersen (1986a) which can be linked to the non-linear least squares routine in BMDP. The program utilizes the Gauss-Newton method for non-linear least squares estimation by modifying the scoring algorithm to accommodate maximum likelihood problems; see Jennrich and Moore (1975) or Petersen (1986b) for details. The parameter estimates and their estimated standard deviations are given in Table 1.

The calculation of the 'residuals' $r_i = -\log [S(t_i; z_i; \hat{\theta}, \hat{\beta})]$ (see, for example, Cox and Oakes (1984), p. 89, or Kalbfleisch and Prentice (1980), p. 96) of model 1 and plotting r against the logarithm of the proportion of residuals exceeding r shows a good fit to a straight line with slope -1 . Thus model 1 provides an adequate description of the data. Nearly all the asymptotic correlations between the parameter estimates of the three models are low. Only the correlation between $\hat{\beta}_{\text{Age}}$ and $\hat{\beta}_0$ is moderate. It is always negative and its lowest value is -0.65 .

From Table 1 we see that the likelihood ratio statistics for models 2 and 3 are highly significant compared with model 1. Therefore, inferences based on these models may be extremely misleading. For example, age has only a significant effect on the exit rate of the first episode, whereas in the second episode its effect is negligible. An inference based on model 3, however, would suggest that age always has a significant effect on the exit rate from unemployment to re-employment. Further examples of this kind can be found in Table 1.

Finally unemployment benefits are not included as a covariate. In the West German unemployment insurance system benefits can be claimed by an employee who has contributed for a certain time: on becoming unemployed such people are entitled to unemployment benefit for 12 months. Using hazard rate models we consider

TABLE 1

Parameter estimates and estimated standard errors for the three models of unemployment duration

	Model 1		Model 2		Model 3
	First spell	Second spell	First spell	Second spell	
Constant β_0	-2.655 (0.0901)	-2.703 (0.1705)	-2.677 (0.0796)		-2.564 (0.0777)
θ	1.504 (0.0207)	1.737 (0.0395)	1.556 (0.0183)		1.489 (0.0179)
<i>Covariates</i>					
Nationality (0 \equiv German; 1 \equiv foreigner)	-0.139 (0.1065)	-0.005 (0.1097)	-0.154 (0.1052)	-0.005 (0.2044)	-0.151 (0.0930)
Sex (0 \equiv male; 1 \equiv female)	-0.577 (0.0610)	-0.505 (0.1239)	-0.598 (0.0600)	-0.437 (0.1242)	-0.624 (0.0542)
Age (years)	-0.017 (0.0023)	-0.001 (0.0041)	-0.019 (0.0021)	0.003 (0.0031)	-0.012 (0.0019)
Limitations arising from ill health (0 \equiv no; 1 \equiv yes)	-0.697 (0.0832)	-0.644 (0.1488)	-0.713 (0.0825)	-0.589 (0.1528)	-0.660 (0.0715)
Occupational status 1 (0 \equiv other; 1 \equiv trained worker)	0.412 (0.0650)	0.326 (0.1098)	0.403 (0.0790)	0.355 (0.1083)	0.395 (0.0555)
Occupational status 2 (0 \equiv other; 1 \equiv white-collar worker)	0.222 (0.0801)	-0.047 (0.1746)	0.215 (0.0789)	-0.003 (0.1776)	0.146 (0.0724)
Duration of first period of unemployment		-0.037 (0.0175)		-0.040 (0.0171)	
Log-likelihood	-6066.1		-6078.0		-6225.4
Likelihood ratio statistic (compared with model 1)			21.8 with 2 degrees of freedom		318.6 with 9 degrees of freedom

conditional duration distributions, given the values of the covariates. Introducing unemployment transfers as a time varying covariate, it follows that for a person receiving unemployment benefit the probability of an unemployment spell of more than 12 months is zero. Therefore, the value 1 of the dummy variable coding 'entitlement to unemployment benefit' is always connected with short durations (less than 12 months) in the sample. This leads to a positive estimate of the corresponding parameter and so people with an existing claim have a significantly increased exit rate compared with other unemployed people (see, for example, Hujer and Schneider (1988)). However, such interpretations and conclusions may be misleading and should be drawn with caution.

Appendix A

Here we derive the likelihood for the general parametric multistate-multiepisode model. The likelihood contribution of individual i is (dropping the subscript i)

$$L = f(t_n, y_n, x_n, \dots, t_1, y_1, x_1 | y_0), \quad (\text{A.1})$$

given that individual i is in state y_0 at time $t_0 = 0$. Using elementary properties of conditional probabilities, equation (A.1) can be written as

$$L = \prod_{k=1}^n f(t_k, y_k | H_{k-1}, x_k) g(x_k | H_{k-1}) \quad \text{where } H_0 = \{y_0\}. \quad (\text{A.2})$$

In the following we only use the first factor on the right-hand side of equation (A.2). If the marginal distribution $g(x_k | H_{k-1})$ of the covariates depends on the relevant parameters, it

becomes necessary to specify a parametric form for the marginal distribution of the covariates. If this is impossible, we can nevertheless use the first factor in equation (A.2) as a partial likelihood, but there is a loss of efficiency.

The last spell of an individual may be censored, and we have

$$L = \prod_{k=1}^n [\lambda_{y_k}^k(t_k | H_{k-1}, x_k)]^{\delta_k} \exp \left[- \int_{t_{k-1}}^{t_k} \lambda^k(u | H_{k-1}, x_k) du \right] \quad (\text{A.3})$$

where $\delta_k = 1$ for $k = 1, \dots, n-1$ and $\delta_n = 0$ if the last spell is censored and $\delta_n = 1$ otherwise.

The complete likelihood for the sample is (reintroducing the subscript i)

$$L = \prod_{i=1}^N \prod_{k=1}^{n_i} [\lambda_{y_{ik}}^k(t_{ik} | z_{ik})]^{\delta_{ik}} \prod_{j=1}^m \exp \left[- \int_{t_{i,k-1}}^{t_{ik}} \lambda_j^k(u | z_{ik}) du \right] \quad (\text{A.4})$$

where the vector z_{ik} includes x_{ik} and all relevant elements of $H_{i,k-1}$.

Introducing the indicator variables

$$\delta_{ilj}^k = \begin{cases} 1 & \text{if the } k\text{th episode of individual } i \text{ starts in state } l \text{ and at} \\ & \text{time } t_{ik} \text{ a transition to state } j \text{ occurs} \\ 0 & \text{otherwise} \end{cases}$$

and

$$\varepsilon_{il}^k = \begin{cases} 1 & \text{if individual } i \text{ experiences at least } k \text{ episodes and } y_{i,k-1} = l \\ 0 & \text{otherwise,} \end{cases}$$

$i = 1, \dots, N, k = 1, \dots, n_i, l, j = 1, \dots, m, l \neq j$, and rearranging terms yields

$$L = \prod_{l,j} \prod_k \prod_{i=1}^N [\lambda_j^k(t_{ik} | z_{ik}, y_{i,k-1} = l)]^{\delta_{ilj}^k} S_{ij}^k(t_{ik} | z_{ik}, y_{i,k-1} = l)^{\varepsilon_{il}^k} \quad (\text{A.5})$$

where

$$\log [S_{ij}^k(t_{ik} | z_{ik}, y_{i,k-1} = l)] = - \int_{t_{i,k-1}}^{t_{ik}} \lambda_j^k(u | z_{ik}, y_{i,k-1} = l) du.$$

Restriction to a specific transition gives equation (2.8).

References

- Andersen, P. K. and Gill, R. D. (1982) Cox's regression model for counting processes: a large sample study. *Ann. Statist.*, **10**, 1100–1120.
- Arminger, G. (1984) Modelltheoretische und methodische Probleme bei der Analyse von Paneldaten mit qualitativen Variablen. *Vierteljahrshft. Dtschn. Inst. Wirtschaftsforsch.*, No. 4, 470–480.
- Blossfeld, H. P., Hamerle, A. and Mayer, K. U. (1986) Ereignisanalyse: Statistische Theorie und Anwendungen in den Wirtschafts- und Sozialwissenschaften. Frankfurt: Campus.
- (1988) *Event History Analysis—Statistical Theory and Applications in the Social Sciences*. Hillsdale: Erlbaum. To be published.
- Borgan, Ø. (1984) Maximum likelihood estimation in parametric counting process models, with applications to censored failure time data. *Scand. J. Statist.*, **11**, 1–16; correction, 275.
- Cox, D. R. (1972) Regression models and life tables (with discussion). *J. R. Statist. Soc. B*, **34**, 187–220.
- (1975) Partial likelihood. *Biometrika*, **62**, 269–276.
- Cox, D. R. and Oakes, D. (1984) *Analysis of Survival Data*. London: Chapman and Hall.
- Hamerle, A. (1985) Counting process models for event history data with multiple episodes. *Discussion Paper* (in German). University of Konstanz.
- (1987) On the incorporation of left censored observations in analysis of survival or duration data. Preprint. University of Konstanz.

- Heckman, J. J. and Singer, B. (1984) Econometric duration analysis. *J. Economet.*, **24**, 63–132.
- Hujer, R. and Schneider, H. (1988) Unemployment duration as a function of individual characteristics and economic trends. In *Event Histories in Life Course Research* (eds K. U. Mayer and N. B. Tuma). Madison.
- Jennrich, R. I. and Moore, R. H. (1975) Maximum likelihood estimation by means of nonlinear least squares. *Proc. Statist. Comput. Sect. Amer. Statist. Ass.*, 57–65.
- Kalbfleisch, J. D. and Prentice, R. L. (1980) *The Statistical Analysis of Failure Time Data*. New York: Wiley.
- Kay, R. (1983) Multistate survival analysis: an application in breast cancer. Preprint. University of Sheffield.
- Lawless, J. F. (1982) *Statistical models and methods for lifetime data*. New York: Wiley.
- Petersen, T. (1986a) Estimating fully parametric hazard rate models with time-dependent covariates: use of maximum likelihood. *Sociol. Meth. Res.*, **14**, 219–246.
- (1986b) Fitting parametric survival models with time-dependent covariates. *Appl. Statist.*, **35**, 281–288.
- Prentice, R. L., Williams, B. J. and Peterson, A. V. (1981) On the regression analysis of multivariate failure time data. *Biometrika*, **68**, 373–379.
- Ridder, G. (1984) The distribution of single-spell duration data. In *Studies in Labor Market Dynamics* (eds G. R. Neumann and N. C. Westergaard-Nielsen). Heidelberg: Springer.
- Tuma, N. B. and Hannan, M. T. (1984) *Social Dynamics: Models and Methods*. New York: Academic Press.
- Voelkel, J. G. and Crowley, J. (1984) Non-parametric inference for a class of Semi-Markov processes with censored observations. *Ann. Statist.*, **12**, 142–160.
- Wong, W. H. (1986) Theory of partial likelihood. *Ann. Statist.*, **14**, 88–123.
- Wu, S. (1982) A Semi-Markov model for survival data with covariates. *Math. Biosci.*, **60**, 197–206.
-